This listing of claims will replace all prior versions, and listings, of claims in the application.

Listing of Claims:

1. (Currently Amended) A method for detecting similar objects in a collection of such objects, the method comprising:

processing a query to produce the collection of objects;

constructing a plurality of hash tables for the collection of objects produced by processing the query; and, for each of two objects:

modifying a previous method for detecting similar objects so that memory requirements are reduced while avoiding false detections approximately as well as in the previous method, wherein the modifying comprises:

———combining four samples of features into seven supersamples;

———compressing each of the seven supersamples to sixteen ~~a number of~~ bits of precision~~, wherein the number of bits of precision is reduced from a number of bits of precision used in the previous method~~; and

———requiring a number of matching supersamples out of the seven supersamples in order to conclude that the two objects are sufficiently similar, wherein the number of matching supersamples is greater than a number of matching supersamples required in the previous method.

2. (Currently Amended) The method of claim 1, wherein requiring the number of matching supersamples comprises requiring at least six of the seven supersamples to match.

3. (Currently Amended) The method of claim 1, wherein requiring the number of matching supersamples comprises requiring at least five of the seven supersamples to match.

4. (Currently Amended) The method of claim 1, wherein requiring the number of matching supersamples comprises requiring all seven supersamples to match.

5-7. (Cancelled)

8. (Currently Amended) The method of claim 1, wherein the objects are documents, and the method is used in association with a search engine query service to determine clusters of query results that are near-duplicate documents.

9. (Original) The method of claim 8, further comprising selecting a single document in each cluster to report.

10. (Currently Amended) The method of claim 9, wherein selecting the single document is by way of a ranking function.

11-13. (Cancelled)

14. (Previously Presented) A method for determining groups of near-duplicate items in a search engine query result, the method comprising constructing a plurality of hash tables for the items in the search engine query result and, for each of two items being compared:

combining four samples of features into each of seven supersamples;
compressing each supersample to 16 bits of precision; and
requiring five of the seven supersamples to match.

15. (Original) The method of claim 14, further comprising selecting a single document in each cluster to report.

16. (Currently Amended) The method of claim 15, wherein selecting the single document is by way of a ranking function.

17. (Currently Amended) A computer-readable storage medium embodying machine instructions implementing a current method for detecting similar objects in a collection of such objects, wherein the current method comprises modification of a previous method for detecting similar objects so that memory requirements are reduced while avoiding false detections approximately as well as in the previous method, the current method comprising:

———processing a query to produce the collection of objects;

———constructing a plurality of hash tables for the collection of objects produced by processing the query; and, for each of two objects,

————————combining four samples of features into each of seven supersamples~~;~~, and

————————compressing each of the seven supersamples to sixteen ~~a number of~~ bits of precision~~, wherein the number of bits of precision is reduced from a number of bits of precision used in the previous method~~; and

———requiring a number of matching supersamples in order to conclude that the two objects are sufficiently similar, wherein the number of matching supersamples is greater than a number of matching supersamples required in the previous method.

18. (Currently Amended) The computer-readable storage medium of claim 17, wherein requiring the number of matching supersamples comprises requiring at least six of the seven supersamples to match.

19. (Currently Amended) The computer-readable storage medium of claim 17, wherein requiring the number of matching supersamples comprises requiring at least five of the seven supersamples to match.

20. (Currently Amended) The computer-readable storage medium of claim 17, wherein requiring the number of matching supersamples comprises requiring all seven supersamples to match.

21. (Cancelled)

22. (Previously Presented) A computer-readable storage medium embodying machine instructions implementing a method for determining groups of near-duplicate items in a search engine query result, the method comprising constructing a plurality of hash tables for the items in the search engine query result and, for each of two items being compared:

combining four samples of features into each of seven supersamples;

compressing each supersample to 16 bits of precision; and

requiring five of the seven supersamples to match.